**Transcriptomic Mapping of the 5-HT Receptor Landscape**

**Roberto De Filippo¹ and Dietmar Schmitz¹²³⁴⁵**

¹ Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health; Neuroscience Research Center, 10117 Berlin, Germany.   
² German Center for Neurodegenerative Diseases (DZNE) Berlin, 10117 Berlin, Germany.   
³ Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität Berlin, and Berlin Institute of Health, Einstein Center for Neuroscience, 10117 Berlin, Germany.   
⁴ Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität Berlin, and Berlin Institute of Health, NeuroCure Cluster of Excellence, 10117 Berlin, Germany.   
⁵ Humboldt-Universität zu Berlin, Bernstein Center for Computational Neuroscience, Philippstr. 13, 10115 Berlin, Germany.

\* Correspondence to: roberto.de-filippo@charite.de

**Keywords:** 5-HT receptors, Transcriptomics, 5-HT, Serotonin.

# **Summary**

Serotonin (5-HT) plays a pivotal role in regulating a wide range of brain functions, including mood, sleep, and cognition. This study presents a comprehensive transcriptomic analysis of 5-HT receptors (Htrs) covering ≈4 million cells across the whole adult mouse brain, utilizing single-cell RNA sequencing (scRNA-seq) data from the Allen Institute. We report on the differential expression patterns of all 14 known Htr subtypes, revealing a wide diversity in their prevalence and distribution across cell classes. Notably, we found that 65.84% of cells transcribe RNA of at least one Htr and often Htrs were found to colocalize. The transcription patterns of Htrs can accurately inform a random forest classifier to identify specific classes and types of neurotransmitter-releasing cells with surprising success. Leveraging a multiplexed error-robust fluorescence in situ hybridization (MERFISH) dataset provided by Harvard University of ≈10 million cells found in a mouse brain, we analyzed the spatial distribution of each Htr confirming previous findings and uncovering novel patterns of transcription at an unprecedented level of detail. Our findings underscore the complexity of the 5-HT system even at the single-cell dimension and provide new insights into the receptor-mediated mechanisms that underpin diverse neural functions and behaviors. To aid the exploration of Htrs transcription in the datasets we provide a custom online visualizer. This tool enables in-depth analysis at various levels of granularity.

# Introduction

5-HT stands as a pivotal neuromodulator within the brain, orchestrating a diverse array of behaviors and physiological processes that include mood regulation {Dayan, 2009 #1401}, memory {Teixeira, 2018 #924}, sleep {Oikonomou, 2019 #347}, feeding {Lee, 2010 #1406} and decision-making {Homberg, 2012 #1407}. This broad spectrum of influence is remarkably orchestrated by a limited number of 5-HT releasing neurons located in the raphe nuclei, which extend their projections across the entire brain {Descarries, 2010 #937}. The intricacy of the 5-HT system is amplified by the extensive repertoire of Htrs, most of them metabotropic receptors linked to different intracellular pathways, resulting in varied and sometimes opposing effects on neuronal excitability {Marin, 2020 #2938}. These receptors are categorized into 14 types spread across 7 families, presenting a complex landscape of 5-HT signaling. To better grasp the impact of 5-HT release, it is crucial to accurately map the expression of each receptor throughout the brain. Historically, techniques such as immunohistochemistry, autoradiography, in situ hybridization, and microarray analysis have been instrumental in shedding light on this question {Vilaró, 2020 #2939; Hawrylycz, 2012 #2940}. These methodologies have provided foundational insights into the brain-wide effects of 5-HT, contributing significantly to our current understanding {Salvan, 2023 #2941}. However, the introduction of single-cell scRNA-seq and MERFISH has revolutionized our ability to dissect gene transcription. Unlike their predecessors, these techniques offer an unparalleled resolution, scalability and depth of profiling, enabling detailed mapping of each Htr's transcription patterns. Such precision enhances our understanding of 5-HT´s functions, setting the stage for more accurate and nuanced insights into its impact on neural circuits. To elucidate the complexities of the distribution of each Htr subtype, we utilized an scRNA-seq dataset provided by the Allen Institute {Yao, 2023 #2828}, incorporating the associated hierarchical clustering of cells at four nested levels. Spatial distribution data was obtained from a MERFISH dataset by Harvard University {Zhang, 2023 #2887}. This approach revealed new patterns of Htrs transcription and set a new benchmark for examining their distribution against the backdrop of existing literature, contributing to our understanding of the complex roles of 5-HT in brain function. To facilitate broader custom exploration of the Htr landscape throughout the brain, we have developed an online visualization tool that enables the seamless exploration of both the scRNA-seq and MERFISH dataset at different levels of granularity, providing a comprehensive view of the 5-HT receptor distribution.

# Results

**Htrs transcription overview**

We analysed the single-cell scRNA-seq dataset provided by the Allen Institute {Yao, 2023 #2828} focusing on the transcription of Htrs genes across approximately 4 million brain cells passing quality control. The scRNA-seq dataset comprehensively encompassed all known 14 Htr subtypes. 65.84% of cells transcribed RNA of at least one Htr. To evaluate transcription we used the same stringent threshold (log(CPM)>3.5) used by the original authors to determine neurotransmitter release {Yao, 2023 #2828}. Prevalence of Htrs, the percentage of cells expressing a receptor, across the entire dataset was considerably different ranging from 0.09% of Htr3b to 34.26% of Htr1f (Figure 1a). RNA of 6 Htr was found in less than 2.5% of the cells (Htr1d, Htr2b, Htr3a, Htr3b, Htr5b, Htr6). On the other hand, RNA of Htr1f, Htr2a and Htr2c was present in at least 1 every 5 cells. Average amount of RNA transcription also varied across receptors (Figure s1a). Interestingly, the variation in amount of RNA shared around half (R²=0.55) of the variability with the prevalence, i.e., genes that were more widespread across cells also exhibited higher transcription rates within individual cells. In addition to differences in prevalence and transcription levels, the distribution of genes across the brain also showed notable variation. This variation is highlighted by comparing the distribution patterns of the Htr1 and Htr2 families, as showcased through uniform manifold approximation and projection (UMAP) analysis (Figure 1b). The UMAP visualization was color-coded according to neighborhood classification. Neighborhoods, characterized by cursory anatomical proximity and molecular signatures such as neurotransmitter-release {Yao, 2023 #2828}, offer a condensed categorization of cell types (Figure s1b, Table 1). When looking at the UMAP distributions of individual Htr, considerable differences were also present within each family of receptors (Figure s2). We analyzed these differences grouping cells by neurotransmitter, neighborhood or class (Figure s1b). The totality of cells were divided in 34 classes (as defined in {Yao, 2023 #2828}). These categorizations divided cells in a highly skewed fashion (Figure s1c). For example, when looking at neurotransmitter-release, 3 groups (Glut, Gaba and unassigned) made up almost the totality of cells (98.47%). Expectedly, the vast majority of cells was classified as excitatory (Glut, 50.79%) and around 1 every 5 cells was found to release GABA (20.62%). All the other neurotransmitter were found in less than 1% of the cells, in particular, 5-HT releasing neurons (Sero) were found in only 0.04% of the cells. Pattern of Htrs transcription across different neurotransmitter groups exhibited a relatively high mean Pearson correlation coefficient (r=0.6±0.03). Sero and cholinergic neurons (Chol) showed the most distinct patterns of transcription with respectively mean r=0.32±0.04 and 0.37±0.05 (Figure 1c). To better evaluate the uniqueness of Htrs RNA transcription per group, and account for differences in amplitude, not captured by simple correlation, we employed a Random Forest Classifier aiming at decoding the grouping variable solely from Htrs' transcription. Overall accuracy of the model in decoding neurotransmitter was 36.61%. Reflecting the correlation analysis, the confusion matrix showed that Sero and Chol were among the groups with higher true positive (TP) rate (Sero=69.86%, Chol=44.99%). Cells not transcribing any neurotranmitter, not exhibiting a low r beforehand, were, nonetheless, identified even more successfully (81.13%). Moreover, Noradrenaline (Nora) and glycine (GABA-Glyc) releasing neurons were identified at considerable levels (Nora=25.81% and GABA-Glyc=44.7%). To understand the contribution of each Htr in each prediction we calculated the mean absolute SHAP (SHapley Additive exPlanations) values for each receptor and neurotransmitter {Lundberg, 2017 #2921; Lundberg, 2020 #2922}. The SHAP values in association with the mean prevalence enabled us to easily understand the defining features of each group. We can appreciate, for example, that the identification of Sero neurons is determined mainly by transcription of Htr1a and Chol neurons by Htr4 and Htr5b. Crucially, absence of transcription can also contribute to the classification, e.g., cells not transcribing any neurotransmitter were identified mainly by absence of any Htr, and Nora neurons detection was guided by the unique absence of Htr4. When looking at different neighborhoods the accuracy of the model was 51.54%. The model could differentiate best the NN-IMN-GC, TH-EPI-Glut and Pallium-Glut groups (NN-IMN-GC=73.71%, TH-EPI-Glut=65.73% and Pallium-Glut=56.89%, Figure s3a). NN-IMN-GC includes all the cells not releasing any neurotransmitter, their classification was therefore expectedly influenced by absence of any Htr. On the other hand, TH-EPI-Glut cells were characterized by the unique combination of high transcription of Htr7 and low transcription of Htr2a and Htr4, Pallium-Glut cells, instead, exhibited relatively low levels of Htr2c and Htr7. Notably, Htr7 and Htr1f seemed to follow opposite gradients across neighborhoods. Across classes, differences in Htrs transcription were even more striking (Figure 1d). 7 groups could be identified with a TP rate >40%: 04 DG-IMN Glut, 05 OB-IMN GABA, 09 CNU-LGE GABA, 18 TH Glut, 22 MB-HB Sero, 25 Pineal Glut, and 32 OEC (Figure s3b). 04 DG-IMN Glut were characterized by high transcription of Htr4 and absence of the usually prevalent Htr2c. Similarly, 05 OB-IMN GABA cells showed virtual absence of Htr2c as well as low Htr4 and high Htr1f transcription; 09 CNU-LGE GABA cells showed high Htr1b and low Htr7/Htr1a; 17 MH-LH Glut exhibited high levels of Htr5b and Htr4; 18 TH Glut showed high levels of Htr7 and virtual absence of Htr4; 22 MB-HB Sero, mirroring the results showed by Sero neurons, were characterized by high levels of Htr1a; at last, 34 Immune cells were identified by absence of any Htr transcription. The exclusive use of Htrs transcription pattern reached an impressive 13.44% accuracy in decoding classes.   
Correlation between Htrs transcription across the totality of cells ranged from -0.03 (Htr1f-Htr3a) to 0.311 (Htr4-Htr2c). Considerable correlation was also found for the Htr7-Htr2c (r=0.264) and Htr1f-Htr2a (r=0.212) pairs (Figure 1e). Interestingly, correlation patterns were not stable across neighborhoods (Figure s4a). For example, Pallium-Glut exhibited a unique negative correlation between Htr4-Htr2a not visible from the analysis of the entire dataset. Of note, TH-EPI-Glut showed the highest absolute correlation across all neighborhoods with r=0.609 between Htr5b-Htr4 and a unique negative correlation between Htr4-Htr7. To explore the underlying causes of the correlations we analyzed colocalization (co-transcription) between Htrs. Across the entire dataset we observed that the most transcribed genes, Htr1f and Htr2c, were regularly transcribed whenever the RNA of any other Htr was detected (Figure 1f). This was a driving factor for correlation. Looking more in detail across neighborhoods, also here we noticed important differences, mainly explainable by differential prevalence of Htrs in each neighborhood. 86.41±1.69% of Htr-transcribing cells exhibited at least 2 Htrs (Figure 1g), therefore, only in a minority of cases a cell was found to transcribe uniquely one Htr (25.78% of the totality of cells). Surprisingly, 7.4% of Htr-transcribing cells transcribed at least 5 Htrs. The extensive transcription of different Htr families within the same cell points at the complexity of the 5-HT system even at the single cell dimension.   
To facilitate an understanding of the downstream cellular effects of 5-HT, we aggregated receptors according to their main intracellular effector. We aggregated Htr1 and Htr5 due to their inhibitory effect (cAMP decrease); Htr4, Htr6 and Htr7 because of the shared downstream effect of increasing cAMP; Htr2 is the only one that causes an Ca2+ increase while Htr3 is the only ionotropic receptor. For each cell we determined the principal pathway activated by 5-HT by analyzing the detected RNA levels for each Htr, grouping them by intracellular effector and selecting the top-ranked. We grouped the results by neighborhood, informed by the differential Htrs' transcription (Figure 1h). Ht3 were present only in a small minority of subpallium inhibitory neurons. In the telencephalon, the absolute majority of both Pallium-Glut and Subpallium-Gaba cells were linked to Htr1/5, and around one quarter of cells featured Htr2 as primary effector. Subcortical cells exhibited a more balanced partition without any absolute majority and a considerable presence of Htr4/6/7. In the following sections we will take a deeper look at Htrs grouped by intracellular effector, We will take advantage of the information provided by the MERFISH dataset of {Zhang, 2023 #2887} regarding 9 Htrs to analyze in detail their spatial distribution.

**Htr1 & Htr5**

Receptors belonging to these two families have an inhibitory effect on the host cell, they are coupled to Gᵢ and cause a downstream decrease of cAMP and activation of GIRK channels {Sharp, 2020 #2888; McCorvy, 2015 #2889}.   
Htr1a agonists, such as buspirone, are used as anxiolytics. Htr1b and Htr1d agonists, like triptans, are effective in treating migraines by causing vasoconstriction of cranial blood vessels.Htr1a RNA have a stable prevalence of ≈10% across neighborhoods in the scRNA-seq dataset, with virtual absence in the TH-EPI-Glut group (Figure 2a). Htr1a co-localized most frequently with Htr1f, Htr2c and Htr2a (Figure 2b) and only in a minority of cases was transcribed alone (<10%). Transcription across classes was highly correlated between the scRNA-seq and MERFISH datasets (Figure 2a) and showed a good correspondence in absolute values, this was the case for the majority of others Htrs. Highest transcription was found in Sero neurons of the mid- and hindbrain (class 22 MB-HB Sero, Figure 2c), nonetheless, cortical excitatory neurons (01 IT-ET Glut), like in the majority of Htrs, contained the highest absolute number of cells transcribing the receptor. Subclasses located in the hippocampus (HPF) contained most of the cortical cells transcribing Htr1a (see online visualizer, 'Overview genes by class'). To pinpoint the spatial location, we first identified in the scRNA-seq dataset the clusters highly enriched with Htr1a RNA with a threshold of 70%, i.e., to be classified as enriched at least 70% of cells in a cluster must express the receptor. Only 6.52% of Htr1a transcribing cells were contained in enriched clusters, pointing at a relatively low importance of this receptor in the clustering algorithm. Looking at the spatial distribution across divisions, the highest prevalence was found in the pallidum (PAL) and HPF (Figure 2d). At a more granular level, 5 of the top 10 structures by prevalence belonged to the raphe nuclei: dorsal nucleus raphe (DR), nucleus raphe obscurus (RO), nucleus raphe pallidus (RPA), nucleus raphe magnus (RM) and superior central nucleus raphe (CS). The high levels of Htr1a transcription in the raphe nuclei is reflection of the high prevalence in Sero neurons outlined beforehand, the raphe nuclei contain the vast majority of Sero neurons of the brain. The hippocampal structure exhibiting the higher prevalence were the medial entorhinal cortex (ENTm) and the area prostata (APr) while the medial septum nucleus (MS) and the diagonal band nucleus (NDB), two structures linked to generation of theta waves {Winson, 1978 #2908} and containing Chol neurons, contributed substantially to the transcription in PAL. Notably, all these results confirms previous reports of Htr1a expression in the raphe {Haj-Dahmane, 1991 #2924;Sprouse, 1987 #2923}, ENTm {Schmitz, 1995 #2925; de Filippo, 2021 #1086} and MS {Kia, 1996 #2926}. Levels of of transcription were stable across the anterior-posterior axis like in most other Htrs (Figure 2e-f).   
Htr1b is involved in social memory persistance in mouse {Wu, 2021 #2945}. Htr1b exhibited a more diverse pattern of transcription across neighborhoods (Figure 3a) ranging from 10 to 30%. Highest prevalence was observed in the MB-HB-Glut-Sero-Dopa group, i.e., glutamatergic, serotonergic and dopaminergic neurons located in midbrain and hindbrain. Colocalization showed a similar pattern compared to Htr1a (Figure 3b), only a minority of cells transcribed Htr1b alone (<10%). Looking at transcription across classes, the 09 CNU-LGE GABA class showed the highest prevalence (58.06%) closely followed by 22 MB-HB Sero (53.73%) (Figure 3c). High transcription in 09 CNU-LGE GABA was in sharp contrast with Htr1a that showed only minimal transcription in this class (1.61%). Also in this case, 01 IT-ET Glut exhibited the highest absolute number of Htr1b expressing cells, specifically, in a subclass of the nucleus of the lateral olfactory tract (NLOT, see online visualizer, 'Overview genes by class'). 17.48% of Htr1b transcribing cells belonged to highly enriched clusters and the striatum (STR) showed an impressive high prevalence with >30% (Figure 3d), in line with previous reports {Maroteaux, 1992 #2928;Pommer, 2021 #2927}. Caudoputamen (CP), Nucleus accumbens (ACB), olfactory tubercle (OT), lateral septal nucleus (LSc) and the parabigeminal nucleus (PBG) all exhibited a prevalence of >20%. DR and RO of the raphe nuclei also exhibited considerable prevalence. Similarly to Htr1a, Htr1b seems to be specifically linked to Sero neurons, in line with this, they have been reported to mediate self-inhibition in these neurons {Hjorth, 1991 #2932}. Distribution across the antero-posterior axes reflected the high prevalence in STR (Figure 3e-f).   
Htr1d was transcribed at a much lower level, never exceeding 7% prevalence in any neighborhood (Figure s5a). It colocalized at highest levels with Htr2c and Htr1f (Figure s5b) and only rarely was transcribed alone (<5%). Similarly to Htr1b, transcription was highest in 09 CNU-LGE GABA and 22 MB-HB Sero (Figure s5c). Notably, 09 CNU-LGE GABA exhibited the highest absolute number of cells surpassing 01 IT-ET Glut. Only a small minority of Htr1d transcribing cells belonged to enriched clusters (2.08%). The paraventricular nucleus of the thalamus (PT and PVT) showed the highest prevalence at only >4% (Figure s5d-e-f).   
Htr1f, surprisingly, showed the highest levels of transcription of all Htrs in the scRNA-seq dataset. Highest prevalence was found in the Pallium and Subpallium groups (Figure 4a), reaching ≈50%. Other groups showed a prevalence of 30-40% with TH-EPI-Glut at ≈20% (Figure 4a). Htr1f was found to colocalize the most with Htr2a and Htr2c (Figure 4b). In 30% of cases Htr1f was the only Htr transcribed in a cell and colocalization decreased linearly with the number of co-transcribed Htrs (Figure 4b). Notably, the slope of the linear regression between values provided by scRNA-seq and MERFISH was significantly lower pointing at a difference in absolute prevalence per class (Figure 4c). The two datasets are, however, still highly correlated, with 66% of shared variability. This was the case also for Htr2a, Htr2c and Htr4. Htr1f was broadly transcribed across almost all classes, including some non-neuronal cells. Pineal gland cells were a notable exception. In absolute numbers, cortical glutamatergic cells showed the highest transcription. Various subclasses located in l5, claustrum (CLA) and HPF exhibited prevalence >50% (see online visualizer, 'Overview genes by class'). Spatial distribution showed a peculiarly asymmetric pattern with transcription concentrated in the most anterior regions. Highest transcription was observed in STR, olfactory areas (OLF) and the cortical subplate (CTXsp) reaching >20% (Figure 4d). Specifically, the highest transcription was observed in nucleus accumbens (ACB) and olfactory tract (OT), similarly to Htr1b. The accessory olfactory bulb (AOB) was the OLF structure with the highest prevalence. CLA and the endopiriform nucleus (EPd), on the other hand, were the CTXsp structure exhibiting the highest prevalence. Interestingly, in the CTXsp, transcription in Glut and Gaba neurons was anticorrelated. High prevalence in Glut neurons corresponded to lower prevalence in Gaba and vice versa. In CLA and EPd HTR1f was transcribed mainly in Glut neurons, while in the amygdala (LA, BLA, BMA) predominantly in Gaba neurons (see online visualizer, 'Overview genes by brain structure'). Isocortex and HPF also exhibited considerable transcription both in excitatory and inhibitory neurons. The amount of RNA transcription per cell was not linear, with a clear peak in the frontal olfactory areas (Figure 4e-f). High transcription of Htr1f in this region was previously observed using immunohistochemistry {Bruinvels, 1994 #2929}. The broad transcription of Htr1f observed in the scRNA-seq dataset across the entire telencephalon is in line with earlier reports {Vila-Pueyo, 2018 #2933}.   
Both Htr5a and Htr5b were not included in the MERFISH dataset, therefore we do not have any direct spatial visualization of their transcription. Htr5a was transcribed at 8-16% prevalence across all neighborhoods (Figure s6a) and colocalized the most with Htr1f, Htr2c and Htr2a (Figure s6b). Transcription was broadly distributed across many classes, although only at lower levels compared to other Htrs (Figure s6c). Only one cluster was considered enriched with Htr5a in the entire scRNA-seq dataset, 3453 PAG-PPN Pax5 Sox21 Gaba. This cluster was located mainly in the midbrain reticular nucleus (RR, Figure s6d-e). Htr5b was transcribed at a much lower level across neighborhoods (Figure s7a), with a maximum of ≈%5 in TH-EPI-Glut. Surprisingly, even if its overall prevalence was much lower than Htr5a, 10 clusters were found to be enriched in Htr5b. This receptor was transcribed at considerable levels only in the 17 MH-LH Glut class (≈50% prevalence). This was reflected by high levels of transcription in the medial habenula (MH, Figure s7d-e), a structure involved in the response to stress and fear {Chou, 2016 #2913;Soria-Gomez, 2015 #2910;Winson, 1978 #2908;Yamaguchi, 2013 #2909}. Some transcription was also evident in the posterior part of the brain, specifically in the inferior olivary complex (IO), driven by a single supertype, 253 IO Fgl2 Glut, and some structures populated by Sero neurons.

**Htr2**

The Htr2 family is mainly linked to Gq/11 and causes depolarization by increasing intracellular Ca2+. HtraA antagonists, such as atypical antipsychotics (e.g., clozapine and risperidone), are used in treating schizophrenia and other psychiatric disorders. Htr2c antagonists are being explored for their potential in treating obesity and metabolic disorders {He, 2022 #2942; Yao, 2021 #2943}. Htr2a, instrumental in mediating the effects of psychedelics {Nichols, 2016 #854}, is found across the brain with highest prevalence in telencephalic neighborhoods, Pallium-Glut and Subpallium-GABA (Figure 5a). Colocalization was highest with Htr1f and Htr2c (Figure 5b). Highest transcription (≈40%) was found in 01 IT-ET Glut, 07 CTX-MGE GABA and 16 HY-MM Glut classes (Figure 5c). Interestingly somatotatin (Sst) neuron belonging to 07 CTX-MGE GABA, while exhibiting a relatively low prevalence at the subclass level, contained various clusters with >70% prevalence {De Filippo, 2024 #2904}. Htr2a was also prevalent across many other classes across the whole brain. 01 IT-ET Glut exhibited by far the highest absolute number of neurons transcribing Htr2a, specifically in subclasses of l5 and CLA, resembling Htr1f (see online visualizer, 'Overview genes by class'). CTXsp showed the highest prevalence, reaching >12% (Figure 5d). Isocortex and STR exhibited both ≈5% prevalence. At a structure level, two structures belonging to the mammillary complex (dorsal premammillary nucleus, PMd and tuberomammillary nucleus,TMd) were in the top ten by prevalence. The mammillary complex has been linked to Alzheimer´s disease {Huang, 2023 #2915}, and memory {Roy, 2017 #2916}. CLA and the EPd showed the highest absolute prevalence. Interestingly, CLA has been proposed to play an important role in mediating the effects of psychedelic compounds {Doss, 2022 #2917}. Prevalence in the STR was driven by the small bed nucleus (BA), a structure important for the integration of limbic and environmental informations {Lebow, 2016 #2931}. Htr2a transcription in CLA and mammillary complex is in line with a previous report in monkey {López-Giménez, 2001 #2930}. Prevalence of Htr2a was highest in frontal regions of the brain, decaying linearly to virtual absence in the cerebellum (Figure 5e-f).   
Htr2b was found only in a minority of neurons and was not included in the MERFISH dataset. No cluster was found to be enriched with Htr2b. Interestingly, neurons belonging to the Pineal Glut class showed the highest prevalence at 7.34% (Figure s8c).   
Htr2c was found at highest prevalence in the MB-HB-Glut-Sero-Dopa and Hy-EA-Glut-Gaba neighborhoods (Figure 6a). Apart from Pallium-Glut, its prevalence was always >40%. Colocalization was highest with Htr1f, Htr4 and Htr7 (Figure 6b). Transcription was broadly distributed across many different classes, especially subcortically (Figure 6c). Many classes exhibited a >60% prevalence. As usual, cortical excitatory neurons exhibited the highest absolute number of cells expressing Htr2c. Some subclasses in OLF, amygdala and retrosplenial cortex (RSP) exhibited >80% prevalence (see online visualizer, 'Overview genes by class'). The majority of cells transcribing Htr2c RNA belonged to enriched clusters. Highest prevalence was found in STR. Similarly to Htr1b, ACB, CP and OT exhibited the highest prevalence (Figure 6d-e-f). Isocortex prevalence derived from the unique transcription in excitatory neurons of the ventral part of the RSP, curiously the area with lowest transcription of Htr1f, otherwise highly prevalent in all other cortical regions. Htr2a was also minimally expressed in this specific area. High prevalence was observed also in excitatory neurons of the anterior olfactory nucleus (AON), piriform area (PIR and PAA) and amygdala (LA and BLA). Htr2c RNA was found across a variety of structures also in the MB (non in Sero neurons), pons (p), medulla (MY) and cerebellum (CB).

**Htr4, Htr6 and Htr7**

These receptors are all connected to Gs {McCorvy, 2015 #2889}, leading to increasing cellular levels of cAMP. Htr4 modulation in HPF has been found to bidirectionally influence memory formation in mice {Teixeira, 2018 #924}. Htr4, similarly to Htr2c, showed highest prevalence (>40%) in the MB-HB-Glut-Sero-Dopa and Hy-EA-Glut-Gaba groups (Figure 7a). It colocalized the most with Htr2c and Htr1f (Figure 7b). Transcription across classes was broadly distributed, with many subcortical classes showing a prevalence >40% (Figure 7c). Highest prevalence was found in the 17 MH-LH Glut class, specifically in the Chol releasing neurons belonging to this class located in TH. In absolute numbers, transcription in excitatory cortical neurons was comparable to other classes but still the highest, driven specifically by subclasses of CA1, CA2, CA3 and subiculum (see online visualizer, 'Overview genes by class'). Spatial distribution exhibited a peculiar pattern with high prevalence in one specific structure of the STR: OT (Figure 7d-e-f). A subclass of interneurons present in OT (060 OT d3 Folh1 Gaba) showed a >98% prevalence. PAL and HPF also exhibited relatively high prevalence (≈10%). Dentate gyrus (DG) granule cells (037 DG Glut) were one of the reasons of the high prevalence in HPF. Excitatory cells of CA2, CA3 and indusium griseum (IG) also transcribed Htr4 RNA (see online visualizer, 'Overview genes by brain structure').   
We do not have MERFISH information about the rarely transcribed Htr6 and no enriched cluster was present in the scRNA-seq dataset. The 09 NU-LGE GABA class exhibited the highest prevalence with 7.73, still, the absolute majority of neurons transcribing the RNA of this gene were excitatory cortical neurons (Figure s9c).   
Conversely, Htr7 was transcribed in >10% of the totality of cells. It reached ≈60% in the TH-EPI Glut group, and considerable amounts (≈40%) in MB, HB and HY groups (Figure 8a). Colocalization was the highest with Htr2c and Htr1f (Figure 8b). Transcription was broadly distributed across classes present in HY, MB and TH (Figure 8c). It colocalized the most with Htr2c, Htr1f and Htr4. Htr7 was broadly transcribed across classes, especially in subcortical structures. Peak prevalence was found in 10 LSX GABA, 16 MY MM Glut and 18 TH Glut with >60% (Figure 8c). Cortical expression in excitatory neurons is driven primarily by subclasses in CA2 and l2 ENT (see online visualizer, 'Overview genes by class'). Htr7 enriched clusters were located mainly in HY and TH (Figure 8d). At a structure level, the parafascicular (PF) and paraventricular nucleus (PVT) of TH showed the highest prevalence (>30%).

**Htr3**

The Htr3 family is the only ionotropic Htr and it causes direct excitation by allowing the influx of cations. Htr3 antagonists, such as ondansetron, are effective antiemetics used to prevent nausea and vomiting.The Htr3a subunit is required for the formation of a functional channel {Maricq, 1991 #2918} and can form functional homopentameric receptors {Walstab, 2010 #2919}. Heteromeric receptors containing Htr3b have an increased channel conductance and different selectivity {Davies, 1999 #2920}. Htr3a is transcribed almost uniquely in the Subpallium-Gaba neighborhood, with a prevalence of ≈8% (Figure 9a), specifically in the 06 CTX-CGE GABA class (Figure 9c). It is one of the few Htr, together with Htr3b and Htr1d, that is not transcribed the most in absolute numbers in 01 IT-ET glut. It colocalizes mainly with Htr2c and Htr7 (Figure 9b). This Htr was mainly transcribed in OLF, CTXsp, HPF and Isocortex (Figure 9d) and is most prevalent in the anterior part of the brain, although, puzzlingly, with slightly lower amount of RNA per cell (Figure 9e-f). Htr3b was not included in the MERFISH dataset, and no cluster was found to be enriched with this receptor. Htr3b was the least transcribed Htr gene in the entire RNAseq dataset. Similarly to Htr3a, its transcription was delimited to the 06 CTX-CGE GABA class (Figure s10c).

# Discussion

In this work we described the main transcriptional features of each Htr across the entire brain, leveraging two datasets provided by {Yao, 2023 #2828} and {Zhang, 2023 #2887}. We found that Htrs RNA is transcribed in around 2 every 3 cells and 6 Htrs were transcribed in >10% of cells, with Htr1f reaching a peak of 34.26%. Htrs patterns of transcription can be used to decode the identity of cells grouped by neurotransmitter, neighborhoods and classes at an above chance level. Surprisingly, it was common to detect multiple Htrs within a single cell. This points at the great complexity of the 5-HT system even at a cellular level. We can recapitulate our results regarding each Htr by summarizing the defining feature of each receptor: Htr1a is transcribed in an important fraction of Sero neurons of the raphe and some HPF excitatory neurons; Htr1b is transcribed in many inhibitory striatal neurons and Sero neurons; Htr1d, similarly to Htr1b, is transcribed in the striatum, although at much lower levels; Htr1f is widely transcribed in telencephalic structures, especially the Isocortex, with a peak in frontal olfactory structures; Htr2a is prevalent in glutamatergic cells of the cortical subplate (CLA and EPd) and the mammillary bodies (TMd, PMd), and hippocampal interneurons; Htr2b is rarely transcribed and is present in some neurons of the pineal gland; Htr2c is broadly transcribed, especially in the STR, excitatory neurons of the amygdala (LA, BLA and BMA) and RSPv, OLF neurons and structures in MB, P, MY and CB; Htr3a and Htr3b are uniquely observed in cortical gabaergic neurons of the 06 CTX-CGE GABA class; Htr4 is transcribed at high levels in the OT, excitatory cells of the hippocampus proper and DG, and Chol neurons of the TH (17 MH-LH Glut); Htr5a is transcribed at low levels with only one enriched cluster in the MB; Htr5b is also transcribed only in few cells, specifically in Chol neurons of the TH; Htr6 does not feature any enriched cluster, some cells in CA3 transcribed this Htr; Htr7 is widely transcribed in subcortical structures, especially in some TH nuclei (PF, PVT, IAD and PT), the mammillary complex (MM and PMd), the lateral septal nucleus (LSv) and the fasciola cinerea of the HPF. Our analysis is in no way exhaustive, and it is limited in scope by the constraints of a traditional scientific article. To bypass this limit and, at the same time, provide the ability to explore the 5-HT transcription landscape at different depths, we provide a custom online visualizer. The visualizer enables the exploration of: Htrs transcription in the MERFISH dataset; the prevalence of each Htr across neighborhoods, class, subclass, supertype and clusters; an overview of Htrs prevalence across classes and subclasses; and an overview of Htrs prevalence across all brain divisions and structures optionally filtered by neurotransmitter release. Our entire analysis pipeline can be easily modified to enable the exploration of different families of genes. Instructions are available in 'Jupyter notebooks structures' in the methods section. One constraint of our study is the indirect characterization of Htrs through the detection of RNA molecules, rather than direct assessment of their presence. However, this potential limitation is mitigated by the fact that mRNA levels are frequently a reliable indicator of receptor expression {Vilaró, 2020 #2939}. Conversely, while mapping receptors directly allows for precise localization, it fails to differentiate between pre- and postsynaptic expression, an important aspect of understanding receptor function and distribution. This lack of specificity becomes particularly problematic, for example, in the context of Sero neurons, which have extensive projections throughout the brain and exhibit diverse autoreceptors. This complexity is underscored both in our findings and in previous research, highlighting the intricate regulatory mechanisms of serotonin neurotransmission {Hjorth, 1991 #2932; Haj-Dahmane, 1991 #2924}. Our exploration of the Htrs landscape represents a substantial advancement, contributing to our understanding of the 5-HT system's role in brain function and behavior.

# Materials and Methods

**Jupyter notebooks structure**

The entire analysis is contained in 2 jupyter notebooks hosted on Github at https://github.com/RobertoDF/Transcriptomics-5-HT. Within the 'Figures' folder, 'Figure\_1.ipynb' and 'Figure\_2.ipynb' notebooks reproduce all figures contained in the paper. To adapt the code for the visualization of different genes it is sufficient to change the 'family\_name' and 'genes\_families' variables in Utils.Settings.py file. Data is downloaded following the instructions provided by the Allen Institute (https://alleninstitute.github.io/abc\_atlas\_access/intro.html). Notebooks to download the scRNA-seq and MERFISH datasets are contained in the 'Load\_Data' folder. To explore the expression of different genes, it is necessary to download the associated expression matrices by changing the selected genes in the 'Download\_RNAseq\_data.ipynb' notebook, this can be achieved by modifying the cells underneath the headings 'Select genes scRNA-seq' and 'Select genes MERFISH'.

**Data preparation**

We loaded the metadata and the precomputed expression matrices for the scRNA-seq dataset relative to all Htr genes (see 'Load\_data/Download\_RNAseq\_data.ipynb'). We also loaded the metadata relative to the 'cluster\_group\_name' residing in originally in a different .csv file ('Find membership df' in 'Figure\_1.ipynb'). This information is referred to as 'membership'. Each of this data structure is a pandas dataframe that can be easily joined together according to the unique cell label index ('joined' dataframe). A different dataframe containing membership information is created ('joined\_with\_membershiop'), this is necessary because some cells belong to multiple 'cluster\_group\_name' or 'neighborhood' and therefore causes a doubling of some rows. We will use the dataframe contianing 'membership information' only in to visualize information relative to 'cluster\_group\_name'.   
The MERFISH dataset was loaded in a similar fashion (see 'Load data MERFISH' in 'Figure\_2.ipynb'). This dataset is split in 4 different dataframes ('Zhuang-ABCA-1', 'Zhuang-ABCA-2', 'Zhuang-ABCA-3' and 'Zhuang-ABCA-4') stored in a dictionary ('cell\_expression'). We concatenated the 4 dataframe in one data structure called 'data\_merfish' using the '.concat()' pandas method. Additionally, we used the spatial informations of each cell belonging to the MERFISH dataset for the registration to the Allen Mouse Brain Common Coordinate Framework (CCF) and , subsequently, we assigned parcellations labels ('CCF registration and parcellation annotation' in 'Figure\_2.ipynb').

**Overview figure visualization**

This figure relies uniquely on the scRNA-seq dataset. In panel A we use a heatmap to visualize both the amount of transcription per cell and the number of cells transcribing each Htr contained in the dataset. In panel B we used the precomputed UMAP coordinates and plot on the color axis information about the most transcribed gene per selected family.   
In panel C and D we plot the percentage of cells transcribing each Htr grouped by neurotransmitter release. We take advantage of the pandas 'Group by' function to concisely perform this computation: after grouping by the selected variable (in this case 'neurotransmitter') we apply a function called 'percentage\_above\_threshold' to compute the percentage of cells within a group expressing a gene above a threshold. The 'percentage\_above\_threshold' function is defined within the 'Utils.Utils.py' file. The threshold is stored in the 'Utils.Settings.py' file ('threshold\_expression'). The confusion matrix is computed within the 'decodddit' function in 'Figure\_1.ipynb'. Here we use a boolean version of the 'joined' dataframe created using the same threshold ('threshold\_expression'). The dataset containing boolean values for gene expression (joined\_boolean) was filtered to include columns of interest, specifically a selector column (sel) and a list of selected genes (selected\_genes). The resulting DataFrame was indexed by the selector column, which represented the target variable (neurotransmitter type), while the remaining columns contained features corresponding to the expression levels of various serotonin receptor genes (Htr). The features for classification were defined as the expression levels of the serotonin receptor genes, and the target variable was the neurotransmitter type. The dataset was divided into training and testing sets using a stratified sampling approach to ensure that the class distribution of the target variable was maintained in both subsets. The test set comprised 5% of the total data. A Random Forest classifier was initialized with parameters set to 10 decision trees, a maximum tree depth of 10, and 30 parallel jobs to leverage computational resources effectively. The classifier was configured to handle class imbalances by adjusting the class weights. The model was then trained using the training data. The trained Random Forest model was used to predict the neurotransmitter types on the test dataset. The performance of the model was evaluated by comparing the predicted labels with the actual labels. The accuracy of the model was calculated and transcribed as a percentage. Additionally, a comprehensive classification report was generated, providing metrics such as precision, recall, and F1-score for each class. A confusion matrix, normalized by the true labels, was also produced to visualize the model's classification performance across different neurotransmitter types. The data manipulation and analysis were conducted using the pandas library. The machine learning model was implemented using the scikit-learn library, specifically the RandomForestClassifier for classification tasks. The evaluation of the model's performance was performed using scikit-learn's accuracy score, classification report, and confusion matrix functions. SHAP (SHapley Additive exPlanations) values were calculated to interpret the feature importance of the Random Forest classifier. An explainer object was created using SHAP's TreeExplainer, which was specifically designed for tree-based models. The explainer was initialized with the trained Random Forest classifier, and the number of parallel jobs was set to 40 to leverage computational resources effectively. The SHAP values were computed for a sample of the feature set (X\_sample). These values indicate the contribution of each feature to the model's predictions.   
In panel E we plot the correlation between transcription of different Htr genes by using the pandas 'corr()' method.   
 To plot the co-localization data of panel F a dictionary named 'coexp' was initialized to store the co-localization results. This dictionary would eventually hold the percentage of co-localization for each pair of genes. A nested loop was employed to iterate through each pair of selected genes, excluding a placeholder category labeled 'Any Htr'. For each target gene and gene to check, the following computations were performed: Co-localization Calculation: For each gene pair, the boolean DataFrame joined\_boolean was used to check whether both genes were transcribed (True) in each sample. This was done using the .all(axis=1) method, which returned True for rows where both genes were transcribed. The sum of these True values indicated the total number of samples where both genes were co-transcribed. Normalization: This sum was then normalized by dividing it by the total number of samples where the target gene was transcribed. This provided the percentage of samples where the gene pair was co-transcribed relative to the expression of the target gene. Storing Results: The computed co-localization percentage for each gene pair was stored in the coexp dictionary with the gene pair as the key. After computing the co-localization percentages for all gene pairs, the results were converted into a pandas DataFrame for further analysis and visualization. The same colocalization was used in the barplots of panel G.   
For panel H we aggregated Htr expression by family by aggregating the expression levels of specific serotonin receptor genes. These genes were grouped into four primary families: Htr1/5: Summing the expression levels of genes Htr1a, Htr1b, Htr1d, Htr1f, Htr5a, and Htr5b. Htr2: Summing the expression levels of genes Htr2a, Htr2b, and Htr2c. Htr4/6/7: Summing the expression levels of genes Htr4, Htr6, and Htr7. Htr3: Summing the expression levels of genes Htr3a and Htr3b. These aggregated values were combined with additional columns representing neuronal classifications (class, subclass, supertype, and cluster\_group\_name). The columns of the resulting DataFrame were labeled accordingly, and a new column ('Primary Htr family') was added. This column identified the primary serotonin receptor family for each entry by determining the family with the highest aggregated expression.

**Receptor figure preparation and visualization**

This figure relies on both the scRNA-seq and MERFISH datasets. In panel A we plot both the prevalence and the average amount of transcription of the selected gene in the two datasets. We excluded from the analysis the 'NN-IMN-GC' neighborhood because of consistently low transcription across all Htr genes. For the visualization of gene expression patterns across different 'neighborhoods', we utilized the Seaborn library in Python to create point plots. Specifically, we employed the 'sns.pointplot' function to illustrate the expression levels of a given gene across various groups. The 'sns.violinplot' function was used to plot violin plots of amount of transcription per group.   
In panel B we used the same co-localization data used in Figure 1 panel F (scRNA-seq dataset), This barplot is a 'sliced' version of that panel focusing on one receptor at the time. To visualize the number of colocalized genes (barplot on the right), we utilized a boolean DataFrame ('joined\_boolean') to filter for selected genes and focus on the expression status of a particular gene. We then calculated the sum of true values (indicating gene transcription) across each row where the specific gene was transcribed. The distribution of these sums was normalized to obtain the percentage of samples exhibiting co-expression of the genes.   
In panel C on the left we repeat the same computation of panel A but using 'class' as grouping variable. On the right, we plotted the raw number of cells transcribing the selected gene across different classes. We first filtered the 'joined' DataFrame to include only rows where the expression level of a specific gene exceeded a defined threshold ('threshold\_expression'). We then counted the occurrences of each class in this filtered dataset. The top 10 classes with the highest counts were selected for visualization. Using Seaborn's barplot function, we created a bar plot to display the distribution of these classes. The y-axis represented the count of occurrences, while the x-axis denoted the different classes.   
In panel D we plotted the prevalence of the selected gene in brain regions at two different hyerarchical levels, 'division' and 'structure'. Here we take advantage of the high-confidence label integration between the 'scRNA-seq' and 'MERFISH dataset' {Zhang, 2023 #2887}. Each cell of the 'MERFISH' dataset is assigned a cell-type label ('class', 'subclass', 'supertype' and 'cluster') from the clustering of the 'scRNA-seq' {Yao, 2023 #2886}.To analyze the expression of specific genes across different brain regions and neuronal clusters, we utilized a multi-step data processing approach. First, we calculated the percentage of cells within each cluster expressing the target gene above a defined threshold, grouping the data by cluster. This allowed us to identify clusters with high gene expression levels in the scRNA-seq. Next, we focused on clusters with significant gene expression, filtering the 'MERFISH' dataset to include only these enriched clusters. We then computed the prevalence of gene expression within these clusters across different parcellation divisions and structures. This was done by normalizing the number of cells expressing the gene in each division or structure by the total number of cells in that division or structure, expressed as a percentage. The results were visualized using bar plots to illustrate the top 10 parcellation divisions and structures with the highest gene expression prevalence. Additionally, we included an inset pie chart to show the proportion of gene expression attributable to the enriched clusters relative to the total gene expression. This pie chart highlighted the contribution of these enriched clusters to the overall expression of the target gene. This pie plot shows that often only a minority of cells transcribing a selected gene belong to enriched clusters. It is possible to visualize the proportion of cells transcribing a gene grouped by area bypassing the 'enriched cluster' computation using the online visualizer ('Overview genes by brain structure', Data source selector='MERFISH').   
Panel E: We calculated the percentage of cells within each cluster that expressed the target gene above a defined threshold, allowing us to identify clusters with enriched gene expression. The number of such enriched clusters was printed for reference. Next, we focused on cells within these enriched clusters and calculated the prevalence of the target gene's expression across different brain sections. This was done by normalizing the number of cells expressing the gene in each section by the total number of cells in that section, expressed as a percentage. The results were plotted using a line plot to illustrate the gene's prevalence across brain sections.   
Panel F: To visualize the expression of a specific gene in the top four brain sections, we implemented a function plot\_4\_best\_sections. This function aimed to identify and plot the sections with the highest gene expression levels. The steps are as follows: Data Preparation: We first prepared the dataset by selecting the relevant brain sections and ensuring that unassigned parcellation divisions were excluded. We merged this dataset with cluster membership information to provide context for the gene expression data.Gene Expression Calculation: The percentage of cells within each cluster expressing the target gene above a defined threshold was calculated.^This allowed us to identify clusters with enriched gene expression.Section Identification: We calculated the prevalence of the target gene's expression in each brain section. Using these prevalence values, we identified the top four sections with the highest gene expression. Peaks in the expression data, spaced adequately apart, were determined using the find\_peaks function. The top four peaks were selected for visualization. Color Assignment: A specific color was assigned to each of the top four sections to differentiate them in the plots. A predefined list of colors was used to ensure consistency and clarity. Plotting: For each of the top four sections, the gene expression data was plotted. The plot\_slice function was used to generate the plots for each section, and the border color of each subplot was set to match the assigned color for the respective section. This helped in visually distinguishing each section. The final figure comprised four subplots, each representing one of the top four brain sections with the highest gene expression levels, providing a clear and comparative visualization of the gene expression patterns across these key sections.

**Data Visualizer**

The visualizer was built in Python using Matplotlib, Holoviews and Panel libraries. It is available as a jupyter notebook ('Figures/INteractive\_vizs.ipynb') and online. The jupyter notebook can be used by following the installation instructions available in https://github.com/RobertoDF/Transcriptomics-5-HT. The visualizer is deployed and accessible online on the Hugging Face portal. It is organized in 4 different tabs: 'Spatial MERFISH', 'Gene by class/subclass/supertype/cluster', 'Overview genes by class' and 'Overview genes by brain structure'. The 'Spatial MERFISH' and 'Overview genes by brain structure' are associated with the MERFISH dataset, remaining tabs are associated with the scRNA-seq dataset. Each tab is associated to different interactive controls and panels. 'Spatial MERFISH': 5 interactive controls enable the selections of different datasets from {Zhang, 2023 #2887}, brain section, gene, class and subclass. The datasets available are 2 coronal (Zhuang-ABCA-1/2) and 2 sagittals (Zhuang-ABCA-3/4). The brain section selector enables the visualization of different slices. The gene selector enables the selection of a specific gene. Class and subclass selector restrict the visualization to selected groups. 6 panels are provided. From top to bottom: lineplot representing the proportion of cells selected out the cells available across the spatial axis associated to each dataset, lineplot representing the amount of transcription across space of the selected gene, lineplot representing the percentage of cells across space in which RNA of the selected gene was detected (threshold set at 0.3), barplot representing the percentage of Htr positive cells in the selected slice grouped by brain structure (number in each bar is the absolute number of cells) and two panels representing the slice selected with the gene transcription on the left and atlas metadata on the right.   
'Gene by class/subclass/supertype/cluster': 2 interactive controls enable the selections of neighborhood group and gene. The neighborhood selector enables the selection of a specific neighborhood. The gene selector enables the selection of a specific gene. For each class of neurons we provide 3 levels of visualization. On top, violinplots representing the gene prevalence by subclass; in the middle, violinplots representing prevalence by supertype and on the bottom barplots representing prevalence by cluster. Each subclass is color-coded according to the panel available for each class.   
'Overview genes by class': 4 interactive controls enable the selections of class, subclass, type of grouping and sorting. The class and subclass selectors enable the selection of a specific class and subclass, respectively. The plot can begrouped at different levels of detail: classes, subclasses, supertypes and even individual clusters (the number of groups that can visualized at the same time is limited by the maximum recursion depth of Holoviews). The plot can be sorted by the group´s alphabetical name or gene expression. Gene prevalence is represented with a heatmap in which the colorbar is updated according to the limits of the current selection. Y axis is populated by the name of the groups selected by the "Group by" selector. X axis shows each Htrs.   
'Overview genes by brain structure': 4 interactive controls enable the selections of dazta soirce, division, neurotransmitter and sorting. The division and neurotransmitter selectors enable the selection of a specific brain division and neurotransmitter, respectively. Gene prevalence is represented with a heatmap in which the colorbar is updated according to the limits of the current selection. Gene prevalence is limited to cluster enriched in the according gene (prevalence within cluster of the gene >70%). The y axis is populated by the brain structures belonging to the currently selected brain division. For each division we can refine our selection by isolating neurons releasing a specific neurotransmitter. X axis shows each Htrs.   
First we identify enriched clusters in the scRNA-seq dataset then we look at the proportion of cells belonging to enriched clusters over the total number of cells per region. This creates a problem in the cases where a 5-HT receptor was not deemed important by the clustering algo, in these cases there might be a really small amount of enriched clusters and most cells will be ignored. To solve this we create a data source selector to enable the possibility to bypass scRNA-seq and look at MERRFISH data only. In this case we look directly at the ratio of cells expressing transcribing each gene over the total number of cells per region.

# Acknowledgements

This study was supported by the German Research Foundation Deutsche Forschungsgemeinschaft (DFG), project 184695641 - SFB 958, project 327654276 - SFB 1315, Germany's Excellence Strategy - Exc-2049-390688087 and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 810580). We thank Willy Schiegel and Tiziano Zito for technical help with cluster computing. We thank members of Schmitz lab for discussion. The authors declare that they have no competing interests.

# Data and materials availability

All the code used to process the dataset is available at https://github.com/RobertoDF/Transcriptomics-5-HT. All figures and text can be reproduced using code present in this repository. Access to the original datasets is provided by the Allen Institute at https://portal.brain-map.org/atlases-and-data/bkp/abc-atlas.

# Contributions

Conceptualization, data curation, formal analysis, investigation, visualization, supervision: RDF. Writing - original draft: RDF. Writing - review & editing: RDF, DS. Funding acquisition & resources: DS.

# Figures

**Figure 1. Overview of Htrs translation.**

(A) Barplot showing absolute number of cells expressing each Htrs, amount of expression is represented in greyscale, no threshold is applied. Inset shows the prevalence of each Htr using a threshold (log(CPM)>3.5). (B) UMAP representation color-coded by neighborhood metadata (left), Htr1 (middle) and Htr2 (right) transcription. (C) Htr expression prevalence in cells grouped by neurotransmitter release (top). Confusion matrix of the multi-label random forest classifier showing true label on y axis and predicted label on x axis (middle). Matrix of absolute SHAP values for each group and receptor (bottom). (D) Htr expression prevalence in cells grouped byclass. (E) Htrs expression correlation matrix. (F) Htrs colocalization matrix. Each dot represents the percentage of colocalization of gene on x axis in cells expressing gene on y axis. (G) Top: Percentage of cells expressing the number of Htrs on the x axis. Percentage of cells expressing the gene on x axis expressing at least another Htrs gene (middle) or at least other 4 Htrs (bottom). (H) Pie charts representing the proportion of principal Htrs grouped by intracellular effector for each neighborhood. Principal effector was identified by summing the expression of Htrs. Each number represents the number of cells in thousands.

**Figure 2. Htr1a transcription**

(A) On the left, dotplot representing Htr1a prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr1a RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr1a RNA (left). Number of Htrs RNA detected in cells expressing Htr1a RNA (right). (C) Prevalence of Htr1a RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr1a RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr1a RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr1a RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr1a RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 3. Htr1b transcription**

(A) On the left, dotplot representing Htr1b prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr1b RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr1b RNA (left). Number of Htrs RNA detected in cells expressing Htr1b RNA (right). (C) Prevalence of Htr1b RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr1b RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr1b RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr1b RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr1b RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 4. Htr1f transcription**

(A) On the left, dotplot representing Htr1f prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr1f RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr1f RNA (left). Number of Htrs RNA detected in cells expressing Htr1f RNA (right). (C) Prevalence of Htr1f RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr1f RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr1f RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr1f RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr1f RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 5. Htr2a transcription**

(A) On the left, dotplot representing Htr2a prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr2a RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr2a RNA (left). Number of Htrs RNA detected in cells expressing Htr2a RNA (right). (C) Prevalence of Htr2a RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr2a RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr2a RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr2a RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr2a RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 6. Htr2c transcription**

(A) On the left, dotplot representing Htr2c prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr2c RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr2c RNA (left). Number of Htrs RNA detected in cells expressing Htr2c RNA (right). (C) Prevalence of Htr2c RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr2c RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr2c RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr2c RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr2c RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 7. Htr4 transcription**

(A) On the left, dotplot representing Htr4 prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr4 RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr4 RNA (left). Number of Htrs RNA detected in cells expressing Htr4 RNA (right). (C) Prevalence of Htr4 RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr4 RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr4 RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr4 RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr4 RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 8. Htr7 transcription**

(A) On the left, dotplot representing Htr7 prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr7 RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr7 RNA (left). Number of Htrs RNA detected in cells expressing Htr7 RNA (right). (C) Prevalence of Htr7 RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr7 RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr7 RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr7 RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr7 RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure 9. Htr3a transcription**

(A) On the left, dotplot representing Htr3a prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, violinplots representing the amount of Htr3a RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr3a RNA (left). Number of Htrs RNA detected in cells expressing Htr3a RNA (right). (C) Prevalence of Htr3a RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr3a RNA by class ranked in descending order (top ten). (D) Ranked prevalence of Htr3a RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr3a RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posterior axis of cells belonging to enriched clusters. (F) Expression of Htr3a RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

# Supplementary Figures

**Figure S1. Htrs transcription and hierarchical clustering visualization.**

(A) Top: Linear regression model fit between prevalence (y axis) and amount of intracellular expression (x axis) for each Htrs. Bottom: violinplots showing distribution of intracellular amount of expression for each Htrs. (B) Heatmaps showing relationship between neurotransmitter, class and neighborhoods. (C) Barplots representing the size of each group when cells are categorized by neurotransmitter (top), neighborhood (middle) and class (botttom).

**Figure S2. UMAP distribution of each Htr**

UMAP-associated neurotransmitter, class and dissection region metadata (first row) and UMAP distribution of each Htr.

**Figure S3. Htrs features across neighborhoods and classes**

(A) Htrs prevalence across neighborhoods (left). Confusion matrix related to the decoding of neighboroods classification from Htrs transcription (middle). SHAP values matrix associated to the neighborhood decoding (right). (B) Confusion matrix related to the decoding of classes classification from Htrs transcription (left). SHAP values matrix associated to the class decoding (right).

**Figure S4. Correlation and colocalization patterns across neighborhoods**

(A) Correlation matrices between pairs of Htrs in each neighborhood. (B) Colocalization matrices of Htrs transcription in each neighborhood.

**Figure S5. Htr1d transcription**

(A) On the left, Htr1d prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr1d RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr1d RNA (left). Number of Htrs RNA detected in cells expressing Htr1d RNA (right). (C) Prevalence of Htr1d RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr1d RNA by class ranked in descending order (top ten). (D) Prevalence of Htr1d RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr1d RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posteroir axis f cells belonging to enriched clusters. (F) Expression of Htr1d RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Supplementary Figure S6. Htr5a transcription**

(A) On the left, Htr5a prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr5a RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr5a RNA (left). Number of Htrs RNA detected in cells expressing Htr5a RNA (right). (C) Prevalence of Htr5a RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr5a RNA by class ranked in descending order (top ten). (D) Prevalence of Htr5a RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr5a RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posteroir axis f cells belonging to enriched clusters.

**Figure S7. Htr5b transcription**

(A) On the left, Htr5b prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr5b RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr5b RNA (left). Number of Htrs RNA detected in cells expressing Htr5b RNA (right). (C) Prevalence of Htr5b RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr5b RNA by class ranked in descending order (top ten). (D) Prevalence of Htr5b RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr5b RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posteroir axis f cells belonging to enriched clusters.

**Supplementary Figure S8. Htr2b transcription**

(A) On the left, Htr2b prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr2b RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr2b RNA (left). Number of Htrs RNA detected in cells expressing Htr2b RNA (right). (C) Prevalence of Htr2b RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr2b RNA by class ranked in descending order (top ten).

**Figure S9. Htr6 transcription**

(A) On the left, Htr6 prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr6 RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr6 RNA (left). Number of Htrs RNA detected in cells expressing Htr6 RNA (right). (C) Prevalence of Htr6 RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr6 RNA by class ranked in descending order (top ten).

**Figure S10. Htr3a transcription**

(A) On the left, Htr3a prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr3a RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr3a RNA (left). Number of Htrs RNA detected in cells expressing Htr3a RNA (right). (C) Prevalence of Htr3a RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr3a RNA by class ranked in descending order (top ten). (D) Prevalence of Htr3a RNA across divisions (left) and structures (right). Inset represents the proportion of cells expressing Htr3a RNA that belongs to enriched clusters. (E) Prevalence (top) and average amount of RNA expression in the MERFISH dataset across the antero-posteroir axis f cells belonging to enriched clusters. (F) Expression of Htr3a RNA detected by MERFISH in 4 representative slices. Border color represents the position on the antero-posterior axis.

**Figure S11. Htr3b transcription**

(A) On the left, Htr3b prevalence across neighborhoods with squared Pearson correlation coefficient (R²) between RNA-seq and MERFISH dataset. On the right, amount of Htr3b RNA detected using RNA-seq (top) and MERFISH (bottom). (B) Amount of colocalization with each Htrs by cells expressing Htr3b RNA (left). Number of Htrs RNA detected in cells expressing Htr3b RNA (right). (C) Prevalence of Htr3b RNA across all classes of cells in RNA-seq and MERFISH dataset. Inset represents the linear regression between the two datasets. On te right, absolute number of cells expressing Htr3b RNA by class ranked in descending order (top ten).

**Table 1. Neighborhoods descriptions**

Description of cells belonging to each neighborhood group.